DOCUMENT RESUME

ED 225 058 CG 016 406

AUTHOR Koitman, David B.; Mayer, Jeffrey P.
TITLE Fidelity and Reinvention in the Implementation of Innovations.
SPONS AGENCY National Science Foundation, Washington, D.C.
PUB DATE Aug 82
GRANT NSF-ISI-7920576-01
NOTE 21p.; Paper presented at the Annual Convention of the American Psychological Association (90th, Washington, DC, August 23-27, 1982).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Adoption (Ideas); Elementary Secondary Education; Evaluation Methods; *Innovation; *Organizational Change; *Program Development; *Program Effectiveness; *Program Evaluation; Social Science Research

ABSTRACT

The field of social innovation policy can presently be seen as divided into two opposing camps: pro-fidelity and pro-adaptation. The former conceptualizes innovations as consisting of a number of relatively well specified components, and argue that rigorously developed, evaluated programs should be implemented with close correspondence to validated models. The latter argues that differing organizational contexts and practitioner needs demand on-site modification. To provide empirical evidence about the pro-fidelity position, seven innovative social programs developed and disseminated nationwide by various public sector organizations (schools, city agencies, prisons, courts) using federal funds were studied. Methods for measuring program fidelity and effectiveness across sites were developed. Results supported the two pro-fidelity assumptions. Four of the seven programs exceeded the acceptable level of fidelity, while the remaining three were in the acceptable range. The second assumption was supported by a significant correlation between fidelity and effectiveness. However, despite the overall support, there was considerable cross-program variation. (Author/JAC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED225058

Fidelity and Reinvention in the Implementation of Innovations

David B. Roitman and Jeffrey P. Mayer

Michigan State University

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

X This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

David B. Roitman

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

This paper was presented at the 90th annual convention of the American Psychological Association, Washington, D.C., 1982. This work was completed under a grant from the National Science Foundation #ISI-7920576-01.

CG 016406

Fidelity and Reinvention in the Implementation of Innovations

The field of social innovation policy research can presently be seen as divided into two opposing camps: "pro-fidelity" vs. "pro-adaptation" researchers (Fullan & Pomfret, 1977). The former conceptualize innovations as consisting of a number of relatively well specified components, and argue that rigorously developed and evaluated programs should be implemented with close correspondence to the validated models or else suffer the consequences of "dilution" (Boruch & Gomez, 1977; Calsyn, Tornatzky, & Dittmar, 1977). Dilution is expected to lead in most cases to reductions in outcome effectiveness. A more moderate fidelity position was taken by Hall and Loucks (1978) who argued that adaptation is acceptable up to the "zone of drastic mutation," beyond which the innovation loses its integrity. Therefore, information concerning the parameters of the drastic mutation zone for any innovation is of crucial importance to policy makers, disseminators, and users.

On the other hand, "pro-adaptation" researchers and practitioners argue that differing organizational contexts and practitioner needs demand on-site modification, virtually without exception (Berman & McLaughlin, 1978; House, et al., 1972). For example, according to Gephart,

> A specific product or procedure is developed for a particular purpose or function....(but)....typically, purposes or functions differ from setting to setting....(and)....although the ideal system would be one which had the needed number and types of components universally required....we seldom know enough in a design effort to create all the component parts (1976, pp. 5-6).

One implication of the pro-adaptation perspective is that the freer users are to adapt programs to their local needs, the more likely that the program which is adopted will last. A second implication is that the more

the program is modified to suit the site, the more likely it is to achieve the outcomes desired by users. A third, even more radical implication of this perspective, is that instead of channeling initial program development funds to specific developer sites, funding should instead be devoted to building the capacities of the local sites to develop innovations independently.

Although the pro-adaptation position has attracted an increasing number of adherents in recent years (Datta, 1981), the research foundations of the principal supporting studies are somewhat tenuous. For example, the widely-cited RAND report on Federal programs supporting educational change (Berman & McLaughlin, 1978) found three dominant patterns for implementation: co-optation (when "the staff adapted the project. . . without any corresponding changes in traditional institutional behavior or practices"), mutual adaptation (when both project and setting were changed) and nonimplementation. The RAND researchers reported that "mutual adaptation was the only process leading to teacher change," and "had a better chance of being effectively implemented" than coopted projects. In addition, they reported a striking absence of high fidelity adoption.

A closer look at the Rand methodology reveals the absence of any bona fide measure of program fidelity. The RAND researchers used as their implementation outcome measure "the extent to which projects met their own goals, different as they might be for each project" (Berman & McLaughlin, 1977, Vol. VII, p. 50). Therefore, their implementation measure was biased to reflect adaptation, rather than fidelity. There was no conclusive way to determine the extent that these programs were modified or what components were changed. Additional doubts concerning the RAND conclusions were raised by Datta (1981), who noted that the "programs" examined were for the most part loosely-defined policy statements, rather than highly specified social programs.

In fact, in a recent article Berman (1980) has advanced the fidelity-adaptation debate considerably by proposing a normative contingency model for implementation strategy. This contingency model implies that different strategies for implementation are most appropriate for different situations (i.e., broad policy statements vs. explicit programs). According to Berman,

> There is no universally best way to implement policy. Either programmed (pro-fidelity) or adaptive implementation can be effective if applied to the appropriate policy situation. ....Policy situations are often so complex that a mix of programmed and adaptive strategies might be more effective than a simple choice between the two.

Berman suggests five situational parameters to be considered when designing an implementation strategy: (1) scope of change (incremental or major); (2) certainty of technology or theory; (3) amount of conflict over policy goals and means; (4) structure of the institutional setting (tightly vs. loosely coupled); and (5) the environment's stability. He argues that relatively structured conditions support the use of programmed (fidelity) approaches, while unstructured situations imply the use of adaptive strategies.

However, despite the clear good sense of Berman's present position, few decision-makers or researchers seem to have adopted it. Instead, as Berman notes, "advocates on both sides seem to be throwing down the gauntlet," (p. 206) and a policy shift on the federal level towards a pro-adaptation position has already begun (Datta, 1981).

The present research was designed in part to provide empirical evidence concerning the viability of the pro-fidelity position. In order to do so, two basic assumptions of this position were examined:

1) Programs which have been operationalized in relatively unambiguous and concrete terms can actually be implemented with acceptable fidelity at adopting sites; and

2) Programs which have been demonstrated to be effective in developer demonstration projects will also produce positive outcomes at adopting sites if implemented with acceptable fidelity.

Methodology and results bearing on these assumptions will be discussed in this paper. Before turning to that discussion, the concept of reinvention will be briefly examined. Future reports will discuss the empirical relationship of reinvention with fidelity and effectiveness.

Reinvention

The term "reinvention" was introduced by Rogers and his colleagues (Eveland, et al., 1977; Rice & Rogers, 1979) to capture the flavor of an active process of change at user sites. "Reinvention" brings to mind the phrase "Not Invented Here," a common slogan used in both public and private sector organizations to describe the rejection of outsiders' ideas simply because they originated outside the organization. Such ideas must be "reinvented" to counter the "Not Invented Here" syndrome.

However, despite the potential usefulness of the term "reinvention," the research by Rogers and his associates may not be generalizable to modified RD&D innovations, since the programs examined by Rogers and his colleagues were disseminated with low component specificity and explicitness. Such programs may behave quite differently from programs which are more well-in-hand (Gephart, 1976). It is therefore fruitful to consider what the concept of reinvention may add to the conceptualization of RD&D innovations. Perhaps the concept of fidelity alone more parsimoniously accounts for the salient phenomena (Taylor, 1980), and reinvention is simply an unnecessary synonym for low-fidelity implementation.

In disagreement with Taylor, and following Larsen and Agarwala-Rogers (1977), we would argue that at least two alternative definitions for reinvention are possible, each of which could distinguish "reinvention" (both conceptually and operationally) from lack-of-fidelity. For example, it is useful to contrast program changes which are relatively creative (adding something new to the program) with relatively uncreative changes. The term "modification" connotes the latter, while the sense of activity and creativity implied by "reinvention" suggests it may best be used to refer to the former type of change.

Note that this conceptualization of reinvention assumes that the program under study was disseminated as a set of relatively concrete components, and that the component set was sufficiently complete to withstand the rigors of transfer without requiring major additions at the adopting site. If these assumptions are not met, the innovation virtually requires "creative adaptation" at the adopting site, and the usefulness of a concept which distinguishes creative from uncreative adaptation is thus questionable.

A second possible definition for reinvention could limit use of the term to instances where changes in the program were made in a deliberate or obvious effort to defend the innovation against the "Not Invented Here" syndrome, by giving the program a character unique to the adopting organization.

These two conceptualizations differ from that of Eveland, et al. (1977); Rice & Rogers (1979); and Larsen & Agarwala-Rogers (1977) who use the term reinvention as a catch-all for "the ways in which an innovation is changed during the process of implementation...(including such notions as) Fidelity...Modification... Expansion... Extent...Flexibility... Complexity...(and) specificity" (Larsen & Agarwala-Rogers, 1977, pp. 136-137). These two potential definitions also differ from the usage of

Rorbaugh and Quinn (1980) who defined reinvention in terms of changing the purpose of an innovation. Noted also that both definitions could describe changes which actually enhance the fidelity of the adopted model. Each of these usages is thus distinct conceptually from "lack-of-fidelity."

Rather than attempting to define the concept a priori, the strategy used in the present study was to collect case study notes on every variation that differed in any way from the variations listed in the fidelity instrument. These qualitative data will be content-analyzed to determine the most comprehensive and meaningful definition of reinvention. Content analysis will also be used to categorize instances of reinvention and determine the frequency of occurrence of different types of reinvention.

Method

Recall that the empirical ground covered by this paper is limited to the examination of the two assumptions stated above, which may be summarized as: (1) Programs can be implemented with acceptable fidelity at adopting sites; and (2) The greater the fidelity, the greater the effectiveness of the implemented program.

These assumptions were examined within the context of the three-phase project described in the first two symposium papers. Briefly reviewing this context, eight programs developed and disseminated using the "modified RD&D approach" were selected. Four of these programs were educational innovations, selected from the catalogue of the National Diffusion Network, Department of Education (Emrick, et al., 1977), and four were criminal justice innovations disseminated by the Exemplary Projects Program of the Justice Department's National Institute of Justice (The National Institute Host Program, 1979). In order to select a subset of the many NDN and NIJ programs for study, the following two criteria were used: (1) potential

for at least 20 site adoptions per program (to provide sufficient statistical power to detect significant relationships); and (2) "organization-wide" quality of each program. This latter criterion was required since the research issues concern organizational rather than individual innovation implementation. Materials for each innovation disseminated by the NDN and the Exemplary Projects Program were independently rated on the selection criteria. Ratings were then discussed by the entire research group, resulting in the program selections. Following the first two phases of the project, one of the four educational programs was dropped from the study, since the data collected in the first two phases had revealed that the program did not adequately meet the original selection criteria. Brief descriptions of the seven programs investigated in the third phase are included in Figure 1.

Sample

Reviewing the sampling strategy described in the previous symposium papers, lists of organizations which had contact with program developers were obtained. Adopters were randomly sampled from these lists. The unit of analysis for Phases Two and Three was the organization in which the program was housed. In some cases, this differed from the organization which made the adoption decision, since implementation entailed creating a new organization or subcontracting to another agency. For example, in one case a crime prevention program was adopted by a police department and later moved to the town's Bureau of Neighborhood Associations. In several other cases, alternative schools were created by district school boards to administer Experienced Based Career Education or Focus programs.

A subsample of the Phase Two respondent organizations were selected for inclusion in Phase Three. Ten organizations from each of the seven

innovations were chosen to be site visited. This resulted in a Phase Three sample size of 70 organizations. Three criteria influenced this selection process. The most important criterion required including organizations that exhibited a range of fidelity scores which were calculated from Phase Two data. Thus, for each innovation, three organizations were selected from above and below one standard deviation from the mean and four were selected from the mid-range. This resulted in ten sites that varied from high to low on fidelity. The second criterion was the length of a program's existence in the organization. If possible, programs that had been in existence for at least two years were selected in order to conform to the requirements of studying routinization. Finally, a broad geographic distribution was sought. In sum, 70 organizations that exhibited variability on fidelity, had been in place for two years, and were located across the entire continental U.S. were selected for site visits.

Measuring Program Fidelity

The five step model for developing a fidelity instrument proposed by Hall and Loucks (1978) was utilized, with several modification to suit the scope and purpose of the study.

Preliminary identification of innovation components. The purpose of the present study involved examining the viability of the pro-fidelity position. Therefore, rather than attaining a comprehensive description of the innovation in practice by interviewing users as well as developers, it was decided to limit the sources for component identification to those individuals who were involved with the program before it had an opportunity to be modified or reinvented at adopting sites. Consequently, the sample of respondents for component identification was limited to several actors at the developer site.

Each developer site was visited by two members of the project staff. Interviews with several staff members at each site were tape-recorded. All written materials and tapes for each innovation were independently content-analyzed by two staff members in order to identify innovation components. The components were selected to conform to the following criteria: (1) The component should be an observable activity, material or facility. If not observable, the implementation of the component should be verifiable through interviews with staff members and clients of the implementing organization. (2) Each component should be logically discrete from other components, and wherever possible, should not depend on the implementation of other components. (3) Each component should be relatively "innovation-specific"; practices which are common to other programs in the organization should not be considered components. (4) The list of components should exhaustively describe the innovation.

Preliminary identification and scaling of variations. The methodology pioneered by Hall and his associates for measuring implementation requires the identification of "variations" for each of the innovation's components. These variations are scaled as "ideal," "acceptable," or "unacceptable." Thus, fidelity is not measured simply by the number of components implemented at the user site, but can be represented by a "fidelity score" which reflects the extent of component variation at the site.

Due to the limited number of adopters of each innovation, researchers who had visited the original innovation sites generated variations (rather than obtaining variations through extensive interviews of adopters). Subsequent modifications of variations were based on in-depth conversations with the innovation developers and pilot interviews with adopters. In generating variations, the researchers attempted to list discrete, observable, and quantifiable alternatives.

Feedback interviews with developers. In order to check the accuracy of the preliminary identification of components and variations, staff members of developer organizations who had been interviewed previously were re-contacted. This second contact involved mailing a list of components and variations to each staff member. The component variations (i.e., each component-specific set of ideal-acceptable-unacceptable variations) generated by the research team were reviewed by developers with the following questions in mind: "Are these variations realistic? Do they describe the possible implementation of the model program completely, or are there other important variations which should be included? Are the researchers correct in their labeling of variations as ideal, acceptable, or unacceptable?"

Feedback from developers concerning the preliminary identification of components and variations was thus obtained, and appropriate modifications and additions were made to the lists. The result was a list of components and scaled variations for each innovation.

Data collection. The procedure for collecting actual data on implementation involved two pairs of researchers traveling to the sites selected for the on-site sample. Each pair visited 35 sites, and spent two days at each site. Data collection consisted of interviews with respondents from several role positions at each site, observations of pertinent activities and facilities (e.g., block-watch meetings, arbitration hearings, juror orientations, interactions among teachers, aides, and students, etc.), and examinations of archival records. In addition, information concerning program effectiveness was collected during site visits as well. Immediately following each site visit, the researchers discussed their notes and arrived at a consensus decision on the fidelity of each component (ideal, acceptable, or unacceptable).

Reliability and validity. Inter-rater reliability was checked at 13 of the 70 site visits (18%), with each pair of researchers conducting one reliability check for each program. (One reliability check was missed due to logistical problems.) During the checks, both researchers interviewed the same respondents, observed the same activities, and examined the same documents. Forms were coded independently, and results were compared. Using the percentage of exact agreement method, an overall reliability of .81 was achieved. At the sites which were not included in the reliability sample, the researchers interviewed, observed, and examined different data. At these sites they also coded the data independently, but discussed their reasons for coding before arriving at a consensus on final coding decisions. At sites which served as reliability checks, this consensus procedure occurred following the determination of the site's reliability.

Following data collection, a rational-empirical scaling method (Jackson, 1970) was used to determine the best sub-scale structure for the fidelity instrument. These scales are discussed in the Results section, below.

The most serious validity issue concerned agreement among different sources of data (e.g., respondents from different organizational roles, observations of behavior and examination of archival records). For most program components, several data sources were used by the researchers in determining their rating of a particular component's fidelity score. Consequently, the extent of agreement between each source and the researcher's consensus decision for each component was felt to reflect the extent to which the researchers were truly capturing the program fidelity. Rather than examining only a sample of these agreements, the percentage agreement between the researchers consensus ratings and the various data sources

for each component were computed for all components on which multiple sources of data were available (7066 out of 9214 total data sources, or 77%, were multiple sources). The overall percentage agreement between these data sources and respondent's ratings was .96.

Measurement of Program Effectiveness

Recall that each of the seven social programs had been evaluated prior to dissemination by the NDN and Exemplary Project Program. The same outcome criteria that were used in these evaluations were also employed in the present study. These criterion measures are listed in Figure 2.

There was a great deal of variation among sites regarding the quality and format of these measures (e.g., some were aggregated in yearly reports, while others were available only in weekly tabulations). There was also variation in the time periods for which data was available; many sites had data for a time period contiguous to or immediately preceding the site visit, but some sites only had year-old data available. In addition, some sites did not have outcome data on all of the criterion indices or had no evidence of program effectiveness available. Given these limitations on the quality of the outcome data, and given the requirement of obtaining effectiveness scores which could be translated into a common metric for analyses across programs, a ranking procedure was used. This involved the two site-visit teams re-pairing so that each new team had first-hand experience with all sites. The teams then reviewed the available archival outcome data for each program, and final decisions were made on the specific indices to be considered as outcome criterion measures for each program. These outcome criteria are listed in Figure 2. The teams then independently ranked sites within programs, resulting in two sets of rankings for 65 sites (outcome data was unavailable for 5 sites). Each of the seven sets of program rankings were correlated to test the procedure's

reliability. An overall rank-order correlation of .90 (Spearman Rho, with Spearman-Brown formula used to correct for the two sets of raters) was obtained, indicating that the procedure was highly reliable.

Results

The first set of data analyses concerned the distributions of fidelity scores among programs and social policy areas. Figure 2 shows the frequency distributions of the raw fidelity average-item scores. Note that four of the seven programs clearly scored in the acceptable range, with a mean fidelity average-item score across sites greater than one (0 = unacceptable, 1 = acceptable, 2 = ideal). Of the remaining three programs, means of .944, .861, and .860 indicate scores close to the acceptable level. It should be noted that of the four scores which exceeded the acceptable value, two were from the educational policy area and two were from the criminal justice policy area.

An analysis of variance was performed to test for differences between program means. Table 1 shows that the differences among program fidelity means were significant ($F=11.45(6,63)$, $p<.001$, $\omega^2=.4723$). Table 1 also shows the results of a Scheffe post-hoc procedure, indicating the between program differences responsible for the overall significant F value. These results must be qualified by the fact that the use of analysis of variance in this situation rests on the following assumptions: (1) The fidelity metric is an interval-level scale, and (2) that the fidelity score for one program is measured on the same scale as that of another program. Since the seven programs are implemented in different organizational contexts (e.g., elementary schools, courts, prisons, city agencies, etc.), and since the fidelity measure is to some extent program-specific (e.g., the number of components per program ranges from 36 to 103), one might argue that seven different measures are actually employed. However, this can be

viewed as a conservative position; a more liberal view would hold that since identical procedures were used for the development of each program's component set, the fidelity measures are fairly comparable across programs.

Standardization of scores is the logical solution to this problem. However, since this process equates means and variances, it defeats the purpose of an analysis of variance. Standardization can, however, be meaningfully used in the types of correlational analyses required to examine the second pro-fidelity assumption, and was thus employed to increase the meaningfulness of cross-program comparisons.

Recall that the second assumption concerned the relationship between program fidelity and program effectiveness. This assumption was examined both across-program and within-program. The across-program analysis involved examining the correlation between standardized fidelity scores and normalized program outcome ranks. The Pearson correlation produced by this analysis was $r=.3797$, $n=65$, $p=.001$, indicating a fairly strong and significant relationship between fidelity and effectiveness across programs. In order to obtain a better estimate of the true relationship between fidelity and effectiveness, this correlation was corrected for attenuation by using the respective reliability estimates. This resulted in a corrected correlation of .4447. The within-program correlations between overall fidelity and effectiveness are listed in Table 2.

In order to further analyze the relationship between fidelity and outcome effectiveness, a rational-empirical scaling procedure was employed to determine the sub-scale structures of each program's component set. The criteria used in constructing sub-scales included maximizing rational interpretability and internal consistency of sub-scales, and minimizing the correlations between sub-scales and the extent to which items from one

sub-scale correlated with other sub-scales. The resulting sub-scales were then correlated with the normalized outcome rankings within each program. Significant sub-scale--effectiveness relationships are reported in Table 3.

Discussion

The results of the present research provide support for two pro-fidelity assumptions: (1) Programs which have been operationalized in relatively unambiguous, complete, and concrete terms can be implemented with acceptable fidelity at adopting sites; and (2) Explicitly operationalized programs will be effective at adopting sites if they are implemented with reasonable fidelity. Regarding the first assumption, four of the seven programs exceeded the acceptable level of fidelity, while the remaining three fell within the acceptable range. The second assumption was supported by the significant correlation of .3797 (65), $p=.001$ (.4447 when corrected for attenuation) between fidelity and effectiveness.

These results contrast with much of the implementation literature which conveys the impression that implementation with fidelity is generally impossible, dangerous, or foolhardy (e.g., Farrar, DeSanctis, & Cohen, 1979). However, despite the overall support given the assumptions, there is considerable across-program variation with regard to both assumptions. A glance at Figure 3 shows a range between .860 and 1.383 for fidelity average-item scores. Table 2 reveals Pearson correlations between overall fidelity scores and effectiveness ranging from .11 to .78.

It is beyond the scope of this paper to explain this variation. However, it should be noted that in addition to what is reported in this symposium, data on several potential predictors of fidelity was collected. These predictors included the extent of contact with program developers,

the extent of institutional support, staff rewards inherent in the program, and compatibility of the program with the adopting organization. And, it should be recalled that this study also collected data on program modifications which did not neatly fit into the fidelity measure. These modifications will be examined to develop an empirically-based definition of the concept of "reinvention." Finally, recall the situational parameters suggested by Berman (1980) as factors affecting the appropriateness of using a fidelity and/or an adaptation-oriented implementation strategy. These parameters could certainly affect fidelity scores. Hopefully, future empirical studies will uncover the relative importance of these parameters.

In sum, variation in fidelity and effectiveness scores might be explained by the variable relationships diagrammed in Figure 4. Note that four of Berman's five situational parameters are included as exogenous variables. (The fifth parameter, conflict over goals and means, overlaps with the predictor Institutional Support.) Note also that fidelity is conceptualized as an intervening variable which moderates the effectiveness of the program (Tornatzky, 1981). It is suggested that future research in this area considers the general approach indicated by this model, and the contingency approach suggested by Berman (1980), as fruitful guides for designing studies of program implementation.

References

Berman, P. Thinking about programmed and adaptive implementation: Matching strategies to situations. In H.M. Ingram & D.E. Mann (Eds.), Why policies succeed or fail. Beverly Hills, CA: Sage Publications, 1980.

Berman, P., & McLaughlin, M.W. Federal programs supporting educational change, Vol. 7: Factors affecting implementation and continuation. April, 1977, The Rand Corporation, Grant No. R-1589/7-HEW, U.S. Office of Education.

Berman, P. & McLaughlin, M.W. Federal programs supporting educational change, Vol. 8: Implementation and sustaining innovations. Final Report, May, 1978, The Rand Corporation, Grant No. R-1589/8-HEW, U.S. Office of Education.

Boruch, R.F. & Gomez, H. Sensitivity, bias, and theory in impact evaluation. Professional Psychology, 1977, 8(4), 411-433.

Calsyn, R., Tornatzky, L.G., & Dittmar, S. Incomplete adoption of an innovation: The case of goal attainment scaling. Evaluation, 1977, 128-130.

Datta, L.E. Damn the experts and full speed ahead: An examination of the study of federal programs supporting educational change, as evidence against directed development and for local problem-solving. Evaluation Review, 1981, 5(1), 5-32.

Emrick, J.A., Peterson, S.M., & Agarwala-Rogers, R. Evaluation of the National Diffusion Network, Vol. 1: Findings and recommendations (SRI Project 4385). Washington, D.C.: U.S. Office of Education, Department of Health, Education, and Welfare, May, 1977.

Eveland, J.D., Rogers, E., & Klepper, C. The innovation process in public organizations: Some elements of a preliminary model. Springfield, VA: NTIS, March, 1977.

Farrar, E., deSanctis, J.E., & Cohen, D.K. Views from below: Implementation research in education. Cambridge, MA: Huron Institute, 1979.

Fullan, M., & Pomfret, A. Research on curriculum and instruction implementation. Review of Educational Research, Spring, 1977, 47(2), 335-397.

Gephart, W.J. Problems in measuring the degree of implementation of an innovation. Paper presented at the annual convention of the American Educational Research Association, April, 1976.

Hall, G.E., & Loucks, S.F. Innovation configurations: Analyzing the adaptation of innovations. Paper presented at the annual meeting of the American Educational Research Association, Toronto, March, 1978.

House, E.R., Kerins, T., & Steele, J.M. A test of the research and development model of change. Educational Administration Quarterly, 1972, 8(1), 1-14.

Jackson, D.N. A sequential strategy for personality scale development. In C. Spielberger (Ed.), Issues in clinical and community psychology. New York: Academic, 1971.

Larsen, J.K., & Agarwala-Rogers, R. Reinvention of innovative ideas: Modified? adopted? none of the above? Evaluation, 1977, 136-140.

The National Institute Host Program. The National Institute Host Program Report, 1979, 1, pp. 1; 11.

Rice, R.E., & Rogers, E.M. Reinvention in the innovation process. Knowledge: Creation, diffusion, utilization, 1980, 1, 499-514.

Rorbaugh, J.W., & Quinn, R.E. Innovation and organizational performance: A study of the implementation and routinization of a new information technology. Grant proposal submitted to National Science Foundation, July 29, 1980.

Taylor, J. Colloquium presented at Michigan State University, December, 1980.

Tornatzky, L.G. Research on implementation: Implications for evaluation practice and evaluation policy. Paper presented at Evaluation Research Society, Austin, TX, October 1, 1981.

Figure 1
Innovative Social Programs Selected for Study

Education

1. HOSTS (Help Our Student to Succeed)--A diagnostic, prescriptive, tutorial reading program for children in grades 2-6. Tutors are community volunteers and cross-age students. The program includes "pulling out" students from their regular classes at least ½ hour per day.

2. EBCE (Experience Based Career Education)--This program provides career experience outside of school at volunteer field sites for the student. Each career site is systematically analyzed for its educational potential. Students' career and academic abilities and interests are systematically assessed. Individualized learning plans which integrate career experiences and academic learning are utilized. Programs typically take students from grades 11-12, although some also accept students from 9-10.

3. FOCUS (Focus Dissemination Project)--A "school within a school" for disaffected junior and senior high school students. All students are required to participate in a support/problem solving group of 8-10 students and one teacher. Behavioral contracting and a governing board with student representatives are important features. Classes in the Focus program involve individualized, self-paced instruction.

Criminal Justice

4. ODOT (One Day/One Trial)--A jury management system that calls in a certain number of potential jurors per day. Potential jurors come in for that day and if not selected to serve in a trial have completed their obligation. Jurors who are selected serve the length of the trial.

5. CAP (Community Arbitration Project)--Juvenile offenders are sent to a formal arbitration hearing run by the court intake division, rather than to courts. Juveniles have the specific consequences of their actions explained to them with parents and victims frequently present at hearings. Youths are then typically given a number of hours of informal supervision usually involving work in the community. Restitution is also frequently required.

6. SCCPP (Seattle Community Crime Prevention Program)--This program is a three phase attack at residential burglary. It involves the setting up of a neighborhood block watch through proactive targeting of neighborhoods, property marking and inventory, and home security inspections.

7. MCPRC (Montgomery County Pre-Release Center)--Involves the setting up of a residential facility separate from the prison. This facility should be in the community from which most of the inmates are drawn. Inmates are encouraged to work so that they will have a job when they are released. Counseling, social awareness instruction, and behavioral contracting are also part of this program.